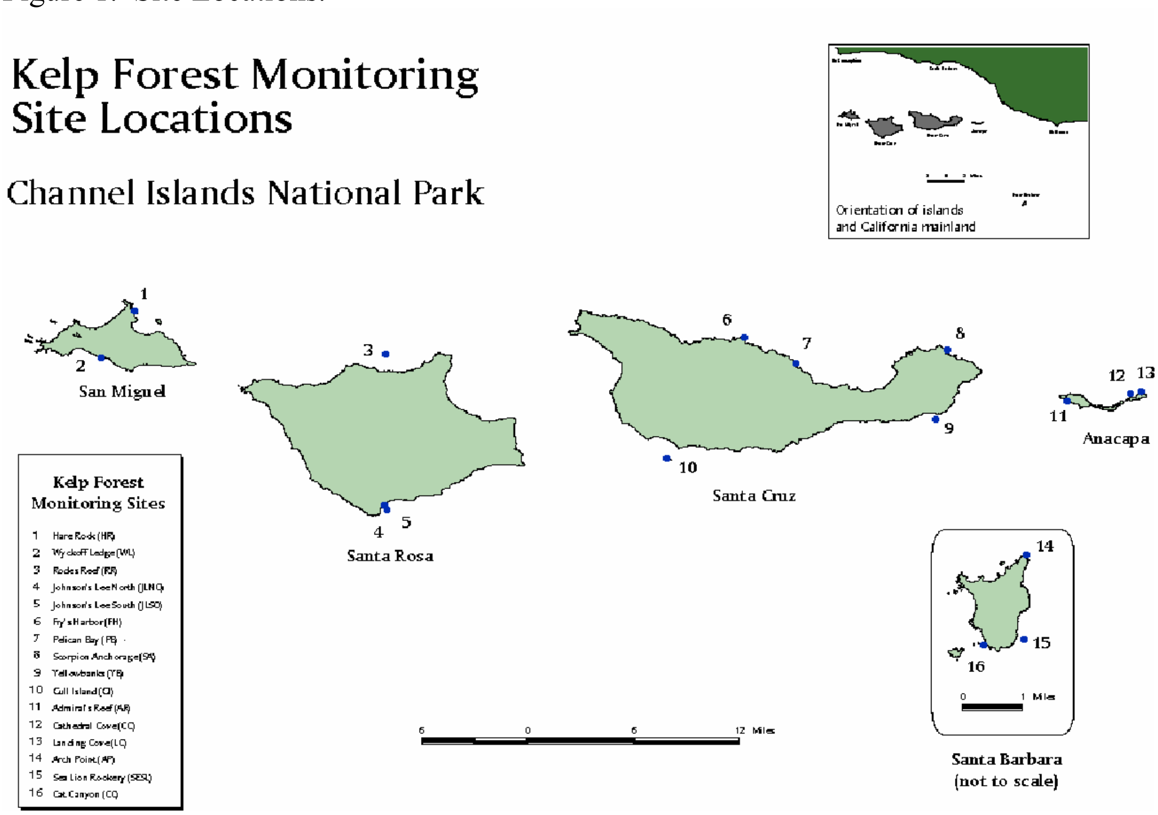
Analysis of Red Urchins *(Strongylocentrotus franciscanus*) Data
From Channel Islands National Park
Paul Geissler
U.S. Geological Survey
Paul_Geissler@usgs.gov, 301-497-5780
DRAFT 3/3/3

The data were collected by Dan Richards and his crew. The use of the data to illustrate analysis methods is greatly appreciated. Information about the survey and protocols are available at http://www.nature.nps.gov/im/units/chis/htmlpages/KFM-HandbookVol1.pdf and http://www.nature.nps.gov/im/units/chis/htmlpages/KFM-HandbookVol2.pdf.

Figure 1. Site Locations.



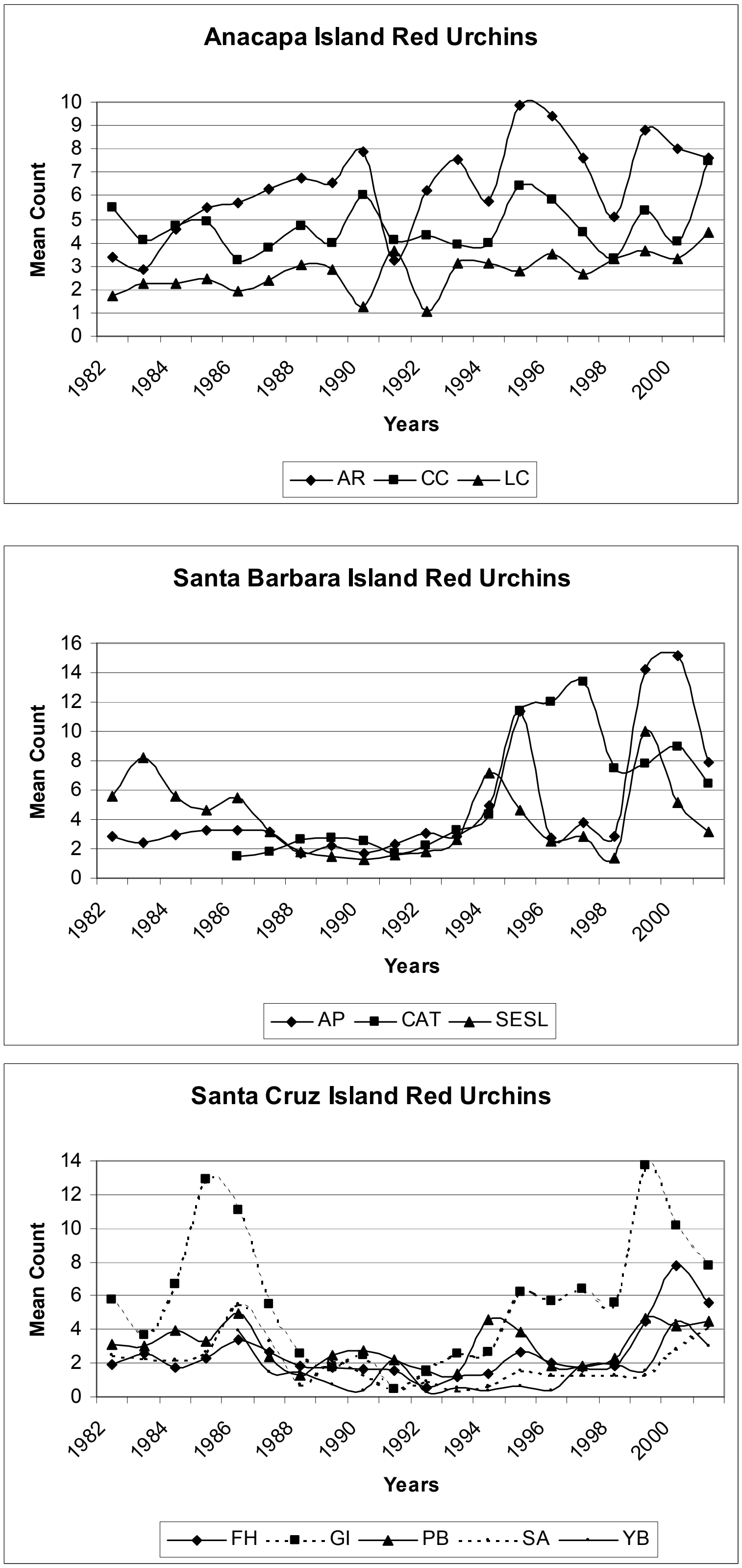
Plots within sites are not true replicates because they were not independently selected (Neter et al. 1996: 1047, Steel and Torrie 1980: 124-126). Consequently, the analysis was conducted on the annual site means (Table 1). There are five islands and several sites on each island. Data were collected since 1982. For the purpose of the analysis, the years have been split into two periods: 1982-1991 and 1992-2001.

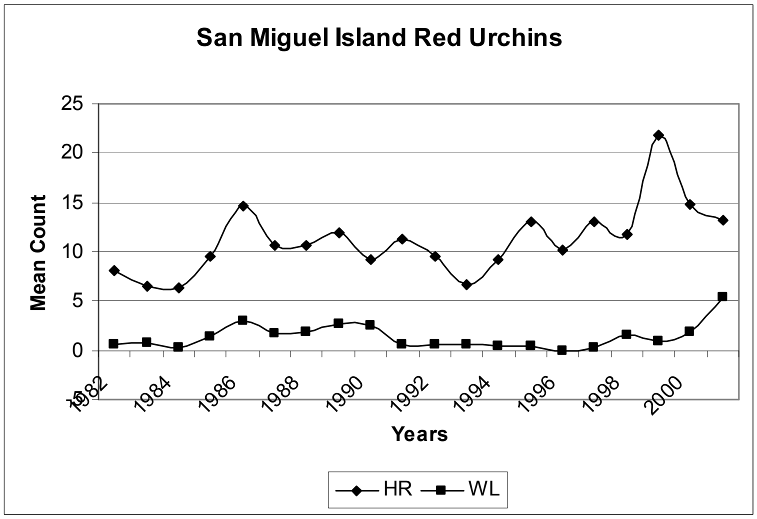
Table 1. Red Urchin Mean Annual Counts for sites at Channel Islands National Park (1982-2001).

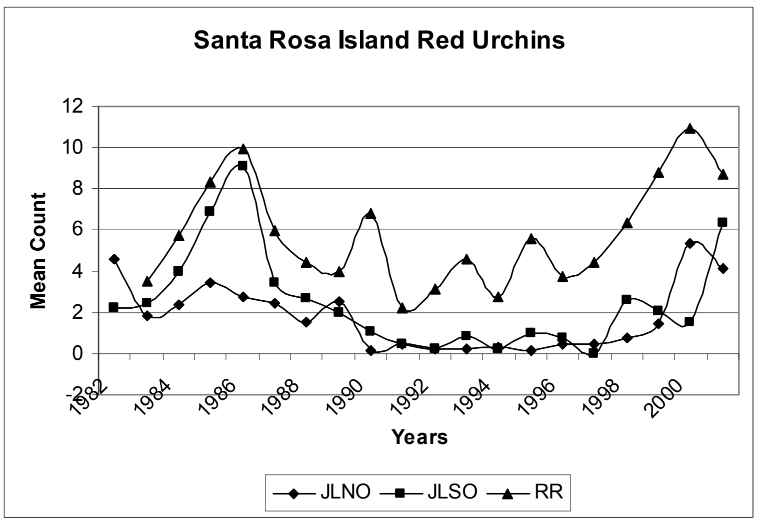
First Period											
Island	**Site**	**1982**	**1983**	**1984**	**1985**	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
AN	AR	3.379	2.875	4.550	5.475	5.725	6.300	6.750	6.550	7.850	3.225
AN	CC	5.500	4.100	4.675	4.925	3.250	3.800	4.700	3.975	6.025	4.125
AN	LC	1.733	2.275	2.225	2.425	1.900	2.375	3.025	2.875	1.250	3.625
SB	AP	2.867	2.375	2.925	3.300	3.250	3.175	1.675	2.175	1.700	2.350
SB	CAT	2.914*	2.734*	2.977*	3.632*	1.450	1.750	2.675	2.775	2.575	1.700
SB	SESL	5.567	8.250	5.575	4.675	5.450	3.200	1.750	1.450	1.225	1.600
SC	FH	1.967	2.525	1.775	2.325	3.350	2.650	1.825	1.750	1.625	1.575
SC	GI	5.800	3.650	6.700	12.875	11.075	5.525	2.550	1.700	2.450	0.500
SC	PB	3.067	3.025	3.900	3.325	4.925	2.350	1.275	2.450	2.725	2.225
SC	SA	2.433	2.325	2.225	2.675	5.450	3.250	0.700	2.125	1.350	0.400
SC	YB	2.683*	2.503*	2.746*	3.401*	3.925	1.500	1.475	0.700	0.325	2.225
SM	HR	8.100	6.450	6.375	9.525	14.675	10.600	10.600	11.850	9.225	11.200
SM	WL	0.567	0.750	0.300	1.450	3.050	1.750	1.825	2.625	2.500	0.600
SR	JLNO	4.575	1.825	2.325	3.400	2.775	2.425	1.525	2.550	0.175	0.450
SR	JLSO	2.200	2.450	3.950	6.875	9.100	3.425	2.700	2.000	1.025	0.475
SR	RR	4.658*	3.525	5.700	8.300	9.950	5.950	4.450	3.950	6.775	2.175
Second Period											
Island	**Site**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**
AN	AR	6.250	7.575	5.775	9.900	9.417	7.583	5.083	8.792	8.000	7.625
AN	CC	4.325	3.925	3.975	6.450	5.833	4.458	3.333	5.333	4.042	7.458
AN	LC	1.075	3.125	3.100	2.750	3.500	2.667	3.292	3.667	3.292	4.417
SB	AP	3.100	2.825	5.000	11.350	2.708	3.750	2.875	14.208	15.208	7.875
SB	CAT	2.225	3.250	4.350	11.400	11.958	13.417	7.500	7.792	9.000	6.375
SB	SESL	1.825	2.675	7.175	4.600	2.542	2.833	1.333	9.958	5.125	3.125
SC	FH	0.550	1.175	1.350	2.650	2.000	1.750	1.833	4.458	7.792	5.625
SC	GI	1.600	2.550	2.625	6.250	5.667	6.417	5.542	13.750	10.167	7.792
SC	PB	1.575	1.350	4.600	3.825	1.792	1.875	2.250	4.708	4.167	4.458
SC	SA	0.900	0.350	0.600	1.575	1.250	1.292	1.250	1.250	2.833	4.083
SC	YB	0.250	0.550	0.400	0.625	0.375	1.792	1.917	1.583	4.458	3.000
SM	HR	9.500	6.575	9.250	13.050	10.083	13.083	11.708	21.792	14.750	13.208
SM	WL	0.575	0.550	0.450	0.500	0.000	0.292	1.542	0.958	1.875	5.417
SR	JLNO	0.200	0.200	0.275	0.150	0.417	0.417	0.792	1.417	5.375	4.083
SR	JLSO	0.225	0.800	0.200	0.975	0.750	0.000	2.583	2.042	1.542	6.333
SR	RR	3.100	4.550	2.750	5.575	3.750	4.458	6.375	8.792	10.958	8.708

* Imputed values

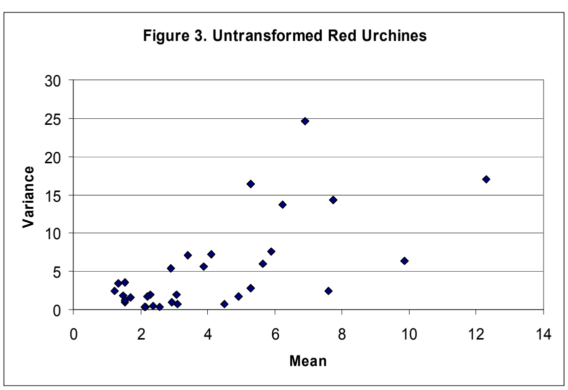
Figure 2. Plots Of Mean Counts By Island.







I checked to see if the data needed to be transformed because of nonnormality or unequal error variances (Neter et al. 1996: 126-134). The variance on the untransformed mean counts increased with their mean (Figure 3). A log transformation [ln (mean count + 0.075] removed that dependence (Figure 4). The logarithm of zero is not defined, so some constant must be added. I used half the lowest value to reduce the effect of this addition. The log transformation also linearized multiplicative trends, but a square root transformation could also be used. This transformation also improved the normality of the means (Figures 5 and 6).



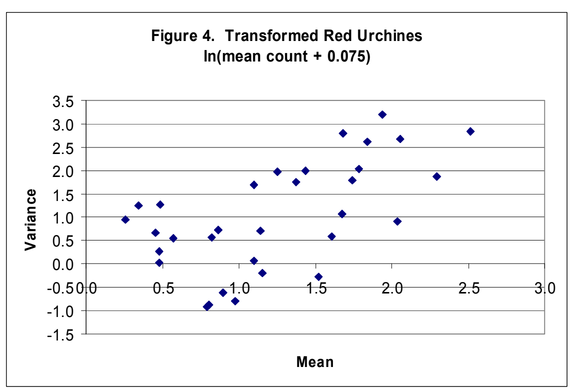


Figure 5. Normal Probability Plot
For Untransformed Means

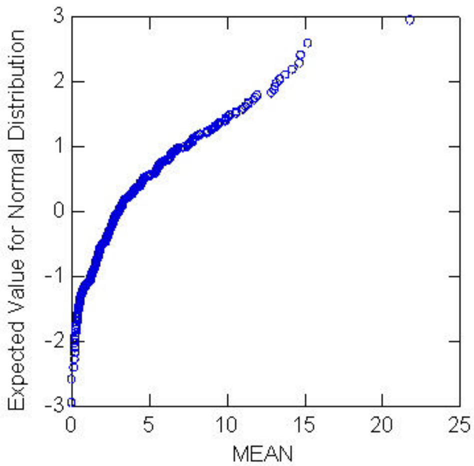
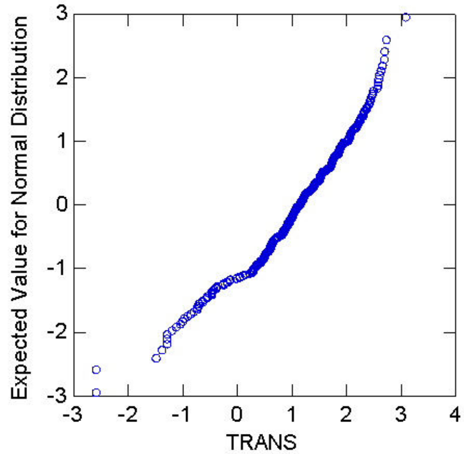


Figure 6. Normal Probability Plot
For Log Transformed Means



Year and site effects are crossed because the same year effect (e.g. weather) applies to all plots and the same plot effects (e.g., location) applies to all years. However, neither plots or year can be rerandomized between years or plots, so this is a repeated measure design, but either years or plots could be considered to be the repeated measurements. This design is variously referred to as a “split-block”, “spilt-plot in time” or a design with both factors in strips (Steel and Torrie 1980:390-400, Cochran and Cox 1957: 306-309).

Within-Subject Functions Approach

The simplest approach is an univariate analysis of within-subject functions (Koch et al. 1988). I calculated the difference between the mean of the second (1992-2001) and the first (1982-1991) periods and the slope of linear regression on years for each site (Table 2). Any function of the annual means counts for a site could be used in this analysis.

Table 2. Difference Between Periods and Slope of Regression on Years for Sites.

Island	Site	Difference	Slope
AN	AR	12.87	0.2219
AN	CC	9.42	0.0475
AN	LC	5.46	0.0917
SB	AP	9.47	0.4243
SB	CAT	9.88	0.6217
SB	SESL	7.99	-0.0442
SC	FH	5.06	0.1285
SC	GI	11.52	0.1060
SC	PB	5.99	0.0220
SC	SA	3.83	-0.0458
SC	YB	3.19	0.0569
SM	HR	22.16	0.3535
SM	WL	2.76	0.0378
SR	JLNO	3.54	-0.0464
SR	JLSO	4.97	-0.1259
SR	RR	11.54	0.1046

An analysis of variance (Table 3) did not detect any differences among islands with the log transformed differences and slopes (Table 2, adding 0.5). Consequently, I analyzed the means of the site differences and slopes. Both the difference and slope showed an increase ($P<0.05$, Table 4).

Table 3. Analysis Of Variance Of Log Transformed Functions From Table 2.

		Log Differences		Log Slope	
Source	**df**	**Mean-Square**	**P**	**Mean-Square**	**P**
Island	4	0.181	0.713	0.115	0.239
Error	11	0.337		0.071	

Table 4. Summary Statistics For Site Functions.

	Difference	Slope	Log Difference	Log Slope
N of cases	16	16	16	16
Mean	8.10	0.122	2.011	-0.516
95% CI Upper	10.76	0.227	2.301	-0.362
95% CI Lower	5.45	0.017	1.721	-0.669
Std. Dev.	4.986	0.196	0.544	0.287
Std. Error	1.25	0.049	0.136	0.072
T	6.503	2.490	14.787	7.167
P	0.000	0.025	0.000	0.000

To look at the power of the survey to detect differences, I calculated the difference that would be detectable for several type I and II error rates (α, β) on the original scale (Cohen 1988: 27-52).

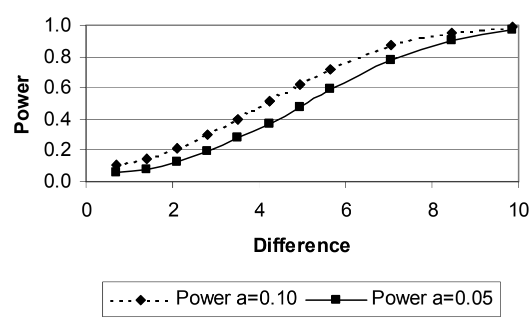


Figure 7. Power To Detect Period Differences

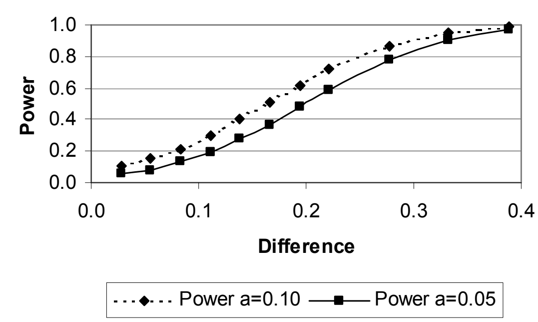


Figure 8. Power To Detect Slope Differences

When one has data available, I think that it is better to estimate the power from the data instead of using a simulation approach (Gibbs et al. 1998), because the estimated power will related specifically to the particular survey and not to some hypothesized situation. Of course when one is planning a survey, data are not available and the simulation approach is useful.

Concern has been expressed about the misuse of power analysis as an alternative to confidence intervals (Hoening and Heisey 2001). All available information about the location of the parameters is contained in confidence intervals. However, confidence intervals are not informative about the ability of a survey to detect future threats and thus about performance of the survey for monitoring. That is the function of the power analysis. Although existing data are used to estimate the power, it is prospective in the sense that one is evaluating the effectiveness of a monitoring survey to detect threats.

Analysis Of Variance Approach – Period Comparison

Another approach is to use an analysis of variance with fixed effects for periods and islands and with random effects for years within periods and for sites within islands. The expected mean squares and F-tests are shown in Table 5 (rules for finding expected mean squares are available in Neter et al. 1996: 1373-1386).

Table 5. Expected Mean Squares

Source	**d.f.**	**Expected Mean Square**	**F**
I: Islands	i-1	$spy\sigma_I^2 + py\sigma_S^2 + \sigma_{SY}^2$	I/S
S: Sites (Island)	i(s-1)	$py\sigma_S^2 + \sigma_{SY}^2$	
P: Periods	p-1	$isy\sigma_P^2 + is\sigma_Y^2 + \sigma_{SY}^2$	P/Y
Y: Years (Period)	p(y-1)	$is\sigma_Y^2 + \sigma_{SY}^2$	
IP: I x P	(a-1)(p-1)	$sy\sigma_{IP}^2 + \sigma_{SY}^2$	IP/SY
SY: S x Y (I, P)	ip(s-1)(y-1)	σ_{SY}^2	

This test for trends (Table 6) is very different from the within-subject functions approach, which uses the consistency over sites of the period differences or slopes to judge the significance. Here the variation among years within periods is used to test for period differences. One is using variation among sites and the other is using variation among years to test for differences.

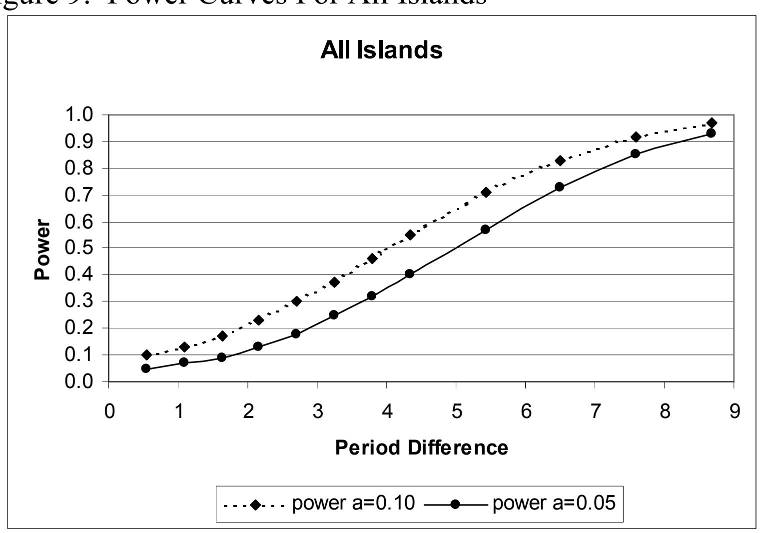
Table 5. Analysis of Variance

		Untransformed		Log Transformed	
Source	**d.f.**	**Mean Square**	**P**	**Mean Square**	**P**
I: Islands	4	60.943	0.715	4.345	0.727
S: Sites (Island)	13	114.822		8.449	
P: Periods	1	51.457	0.203	0.005	0.966
Y: Years (Period)	18	29.426		2.61	
IP: I x P	4	32.702	0.000	2.509	0.000
SY: S x Y (I, P)	225	3.305		0.277	

I used SYSTAT for the analysis, which cannot calculate type 4 sums of squares to allow for missing cells. To calculate the IP and SY (but not other) sums of squares, I imputed the missing values as the mean of the row and column means for a period in Table 1 (Steel and Torrie 1980: 209-214). The degrees of freedom for SY were reduced by the number of missing cells. SAS can calculate these sums of squares without imputation and provides a much better analysis.

The power for an analysis of variance (Figure 9) can be obtained from the table 8.3.1 in Cohen (1988: 273-380). The difference between two means d = (mean 2 – mean 1) / (square root of denominator mean square) = 2 f, u = 1 = numerator degrees of freedom in the F ratio, and n = [(denominator degrees of freedom)/(u+1)] + 1. I used the original scale for these calculations.

Figure 9. Power Curves For All Islands



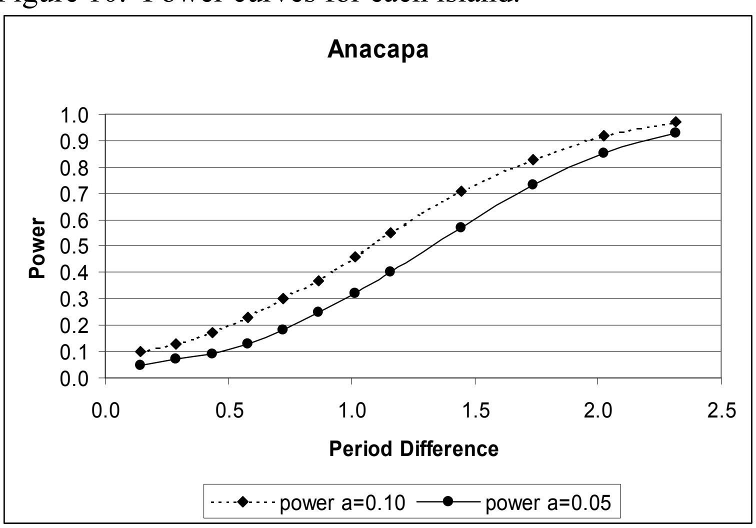
The significant island by period interaction indicates that a separate analysis should be conducted for each island because the period effects differ among islands (Table 7).

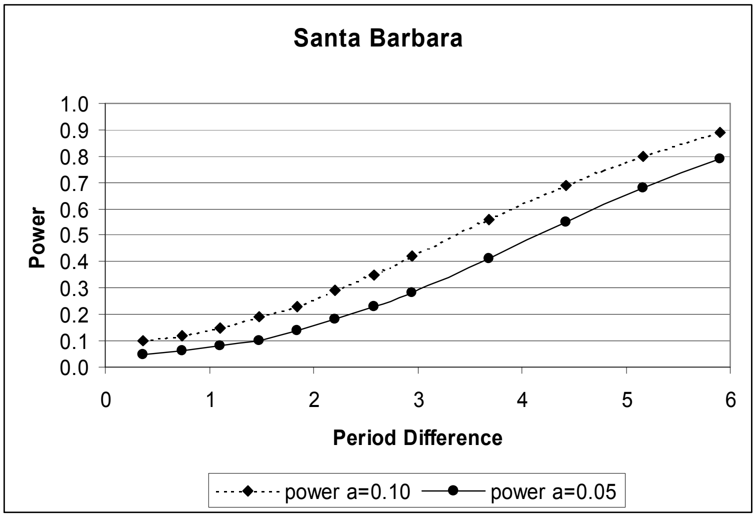
Table 7. Analysis of Variance For Each Island

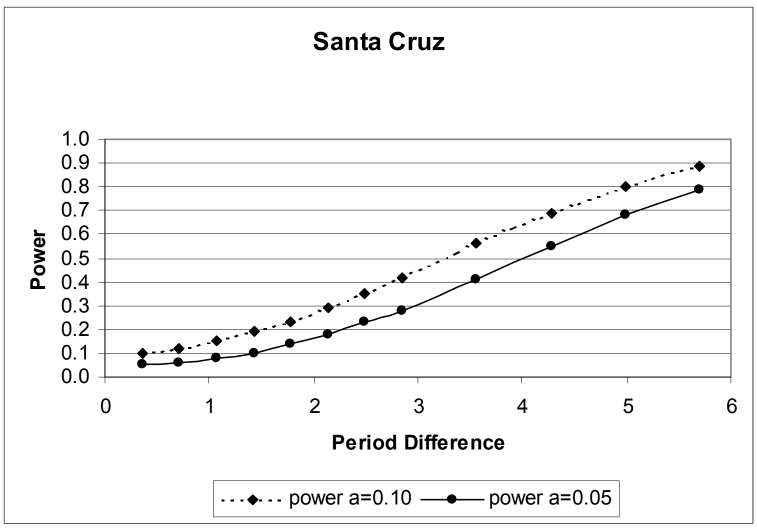
		Untransformed		Transformed	
Source	**df**	**Mean Square**	**P**	**Mean Square**	**P**
Anacapa					
S: Sites	2	68.723		3.676	
P: Periods	1	19.9	0.006	0.837	0.008
Y: Years (Period)	18	2.092		0.093	
Santa Barbara					
S: Sites	2	6.541		0.19	
P: Periods	1	158.903	0.005	5.729	0.005
Y: Years (Period)	13	13.584		0.51	
Santa Cruz					
S: Sites	5	41.617		3.451	
P: Periods	1	0.24	0.893	0.255	0.689
Y: Years (Period)	13	12.697		1.522	
San Miguel					
S: Sites	1	941.143		56.229	
P: Periods	1	11.174	0.262	0.308	0.493
Y: Years (Period)	18	8.318		0.629	
Santa Rosa					
S: Sites	3	59.282		8.968	
P: Periods	1	9.664	0.353	5.905	0.080
Y: Years (Period)	17	10.616		1.699	

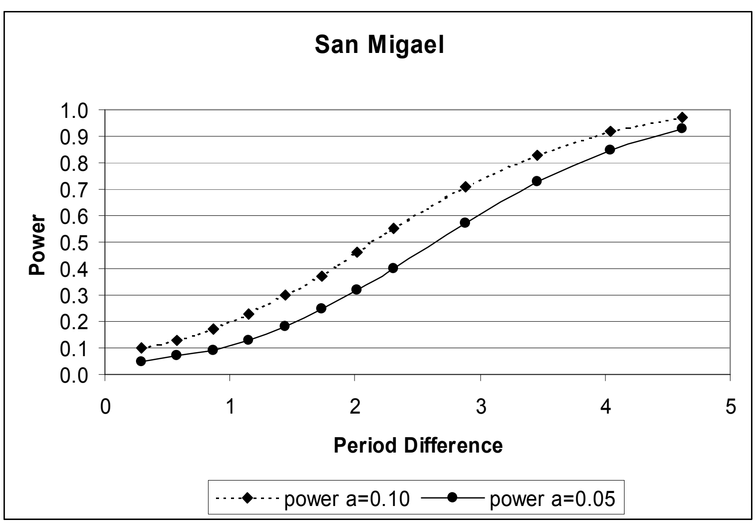
This shows increases only for Anacapa and Santa Barbara ($P<0.05$).

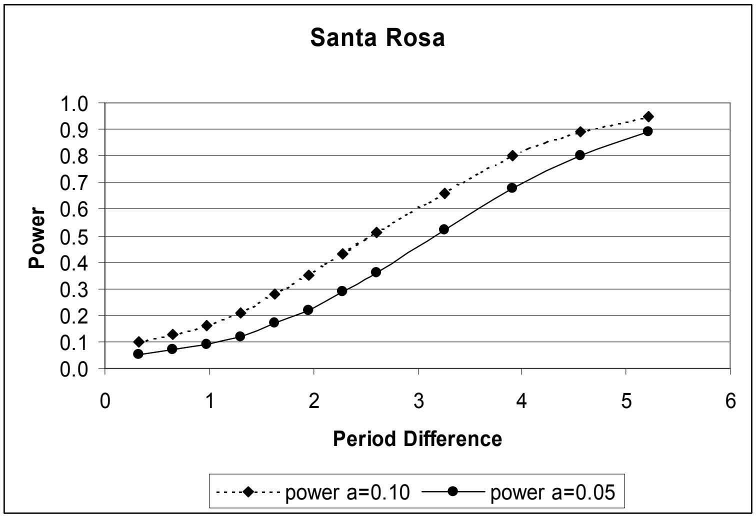
Figure 10. Power curves for each island.




Santa Barbara
Power
Period Difference
power a=0.10
power a=0.05


Santa Cruz
Power
Period Difference
power a=0.10
power a=0.05


San Migael
Power
Period Difference
power a=0.10
power a=0.05



Analysis Of Variance Approach – Slope

Instead of comparing the means of two periods, one can estimate the slope of a linear regression on years. I used the general linear test approach (Neter et al. 1996: 78-80), fitting a full model with categorical years and a reduced model with numeric years (Table 8). The indented effects are partitions of the above effects. I fitted main effects before the interactions, so the main effects would not be adjusted for the interactions.

Table 8. Analysis of Variance

		Untransformed			**Log Transformed**		
Source	**df**	**Sum of Squares**	**Mean Square**	**P**	**Sum of Squares**	**Mean Square**	**P**
Islands	4	243.772	60.943		17.378	4.345	
Sites (Island)	13	1492.68	114.822		109.839	8.449	
Years	19	584.632	30.770		46.982	2.473	
- Slope	1	118.698	118.698	0.046	1.384	1.384	0.469
- Residual	18	465.935	25.885		45.598	2.533	
Islands * Years	76	460.163	6.055		460.163	6.055	
- Slopes	4	98.527	24.632	0.001	98.527	24.632	0.000
- Residual	72	361.636	5.023		38.687	0.537	

This again shows that the island slopes differ (P<0.01) and that islands should be analyzed individually (Table 9).

Table 9. Analysis of Variance For Each Island

		Untransformed			Transformed		
Source	**df**	**Sum of Squares**	**Mean Square**	**P**	**Sum of Squares**	**Mean Square**	**P**
Anacapa							
Sites	2	137.445	68.723		7.351	3.676	
Years	19	57.554	3.029		2.520	0.133	
- Slope	1	28.911	28.911	0.000	1.349	1.349	0.000
- Residual	18	28.643	1.591		1.171	0.065	
Slope Est.		0.120			0.026		
Santa Barbara							
Sites	2	23.206	11.603		0.772	0.386	
Years	19	398.618	20.980		15.146	0.797	
- Slope	1	129.003	129.003	0.009	3.899	3.899	0.022
- Residual	18	269.615	14.979		11.247	0.625	
Slope Est.		0.278			0.048		
Santa Cruz							
Sites	5	206.289	41.258		18.054	3.611	
Years	19	225.438	11.865		25.999	1.368	
- Slope	1	9.123	9.123	0.395	0.351	0.351	0.626
- Residual	18	216.315	12.018		25.648	1.425	
Slope Est.		0.056			0.011		
San Miguel							
Sites	1	941.143	941.143		56.229	56.229	
Years	19	160.901	8.468		11.634	0.612	
- Slope	1	50.897	50.897	0.010	0.297	0.297	0.501
- Residual	18	110.004	6.111		11.337	0.630	
Slope Est.		0.196			0.015		
Santa Rosa							
Sites	3	180.617	60.206		27.526	9.175	
Years	19	202.285	10.647		37.135	1.954	
- Slope	1	1.255	1.255	0.741	2.464	2.464	0.273
- Residual	18	201.030	11.168		34.671	1.926	
Slope Est.		-0.026			-0.036		

Figure 11. Power Curves For All Islands Slope.

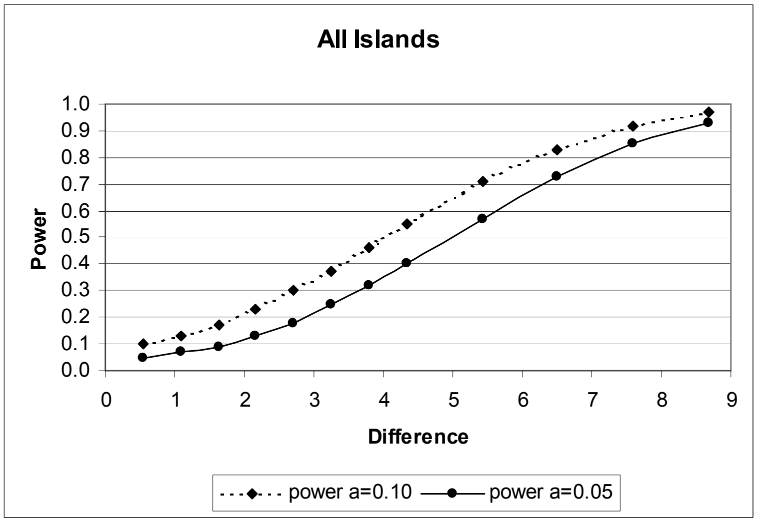
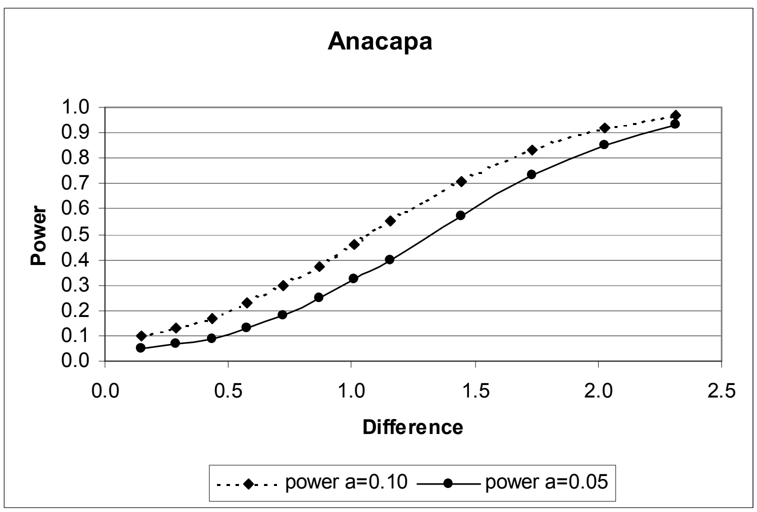
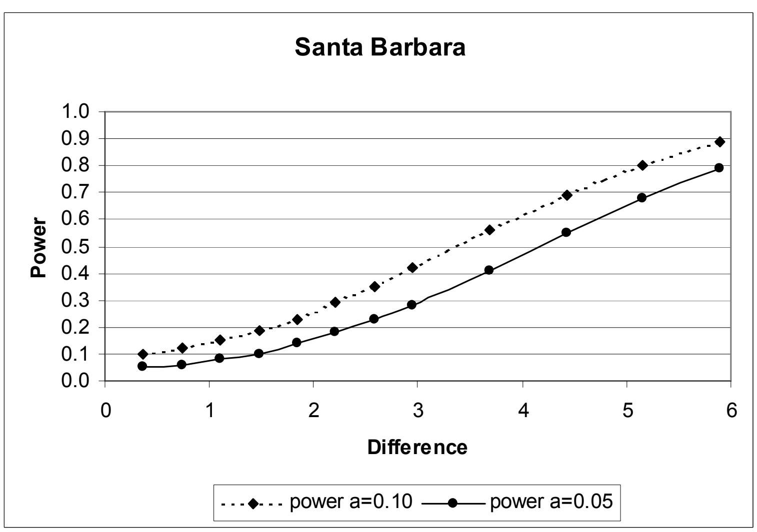
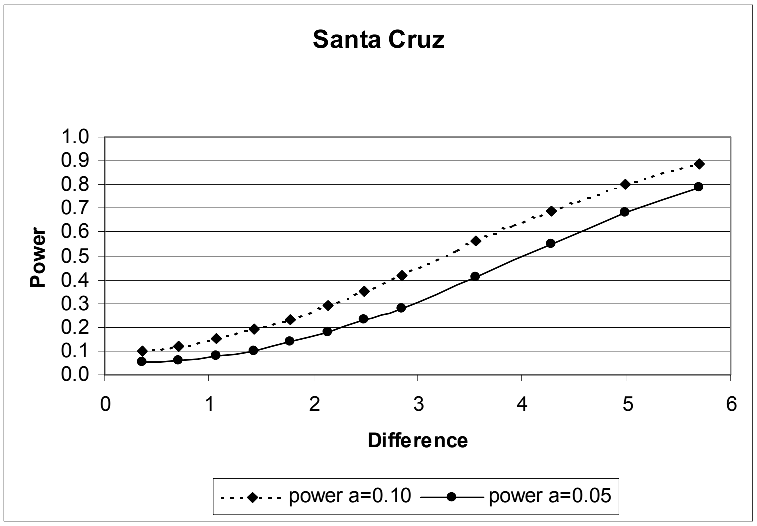
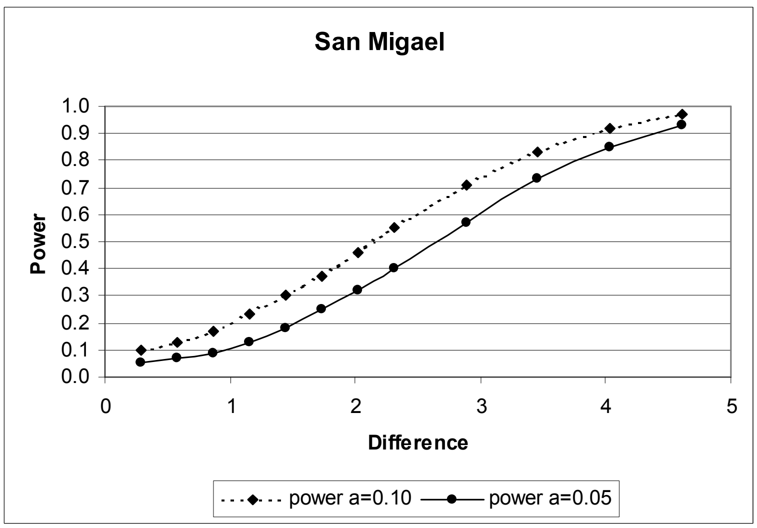


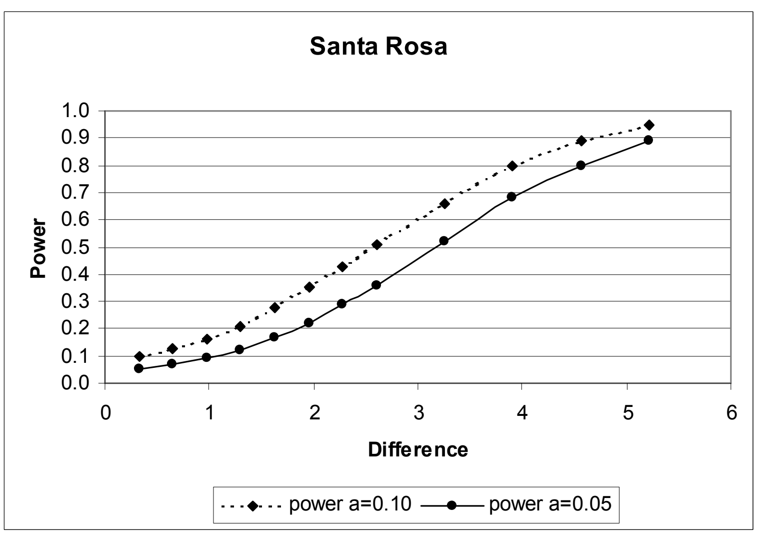
Figure 12. Power Curves For Each Island Slopes.






Santa Cruz
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Power
0
1
2
3
4
5
6
Difference
power a=0.10
power a=0.05


San Migael
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Power
0
1
2
3
4
5
Difference
power a=0.10
power a=0.05


Santa Rosa
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Power
0
1
2
3
4
5
6
Difference
power a=0.10
power a=0.05

References:

Cochran, W. G. and G. M. Cox. 1957. Experimental designs. Wiley, New York

Cohen, J. 1988. Statistical power analysis for the behavioral sciences. Lawrence Erlbaum, Hillsdale, NJ.

Gibbs, J. P., S Droege, and P Eagle. 1998. Monitoring populations of plants and amimals. Bioscience 48(11) 935-940.

Hoening, J.M. and D.M. Heisey. 2001. The abuse of power: the pervasive fallacy of power calculations for data analysis. American Statistician 55(1):19-24.

Koch, G. G., J. D. Elashoff and I. A. Amara. 1988. Repeated Measurements – Design and Analysis in S. Kotz and N. L. Johnson, eds. Encyclopedia of Statistical Sciences, Vol. 8. Wiley, New York

Neter, J, M. H. Kutner, C. J. Nachtsheim and W. Sasserman. 1996. Applied linear statistical models. Irwin, Chicago.

Steel, R. G. D. and J. H. Torrie. 1980. Principles and procedures of Statistics. McGraw-Hill, New York.